RC23783 (W0511-049) November 7, 2005
Mathematics

# IBM Research Report

## Some Theory and Practical Uses of Trimmed L-moments

**J. R. M. Hosking**
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

# Some theory and practical uses of trimmed *L*-moments

J. R. M. Hosking

*IBM Research Division, T. J. Watson Research Center,*
*P.O. Box 218, Yorktown Heights, NY 10598*

**Abstract.** Trimmed $L$-moments, defined by Elamir and Seheult (*Comput. Statsist. Data Anal.*, 2003), summarize the shape of probability distributions or data samples in a way that remains viable for heavy-tailed distributions, even those for which the mean may not exist. We derive some further theoretical results concerning trimmed $L$-moments: a relation with the expansion of the quantile function as a weighted sum of Jacobi polynomials; the bounds that must be satisfied by trimmed $L$-moments; and recurrences between trimmed $L$-moments with different degrees of trimming. We also give examples how trimmed $L$-moments can be used, analogously to $L$-moments, in the analysis of heavy-tailed data. Examples include identification of distributions using a trimmed $L$-moment ratio diagram, shape parameter estimation for the generalized Pareto distribution, and fitting generalized Pareto distributions to a heavy-tailed data sample of computer network traffic.

**Keywords**: Distribution theory, generalized Pareto distribution, $L$-moment ratio diagram.

## 1. Introduction

$L$-moments are measures of the location, scale and shape of probability distributions. They are analogous to the conventional moments but can be estimated by linear combinations of order statistics. $L$-moments are related to expected values of order statistics. Let $X$ be a random variable and let $X_{j:n}$ denote an order statistic, a random variable distributed as the $j$th smallest element of a random sample drawn from the distribution of $X$. Hosking (1990) defined the $L$-moments of $X$ to be the quantities

$$\lambda_1 = \mathrm{E}(X_{1:1}),$$

$$\lambda_2 = \tfrac{1}{2}\,\mathrm{E}(X_{2:2} - X_{1:2}),$$

$$\lambda_3 = \tfrac{1}{3}\,\mathrm{E}(X_{3:3} - 2X_{2:3} + X_{1:3}),$$

$$\lambda_4 = \tfrac{1}{4}\,\mathrm{E}(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}),$$

and in general

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathrm{E}(X_{r-j\,:\,r})\,. \tag{1}$$

Sample estimators of $L$-moments can be used as summary statistics for data samples, and to identify probability distributions and fit them to data. These are similar to the uses of conventional moments, but $L$-moments have the advantage that they exist whenever the mean of the distribution exists, even though some higher moments may not exist, and are relatively robust to the effects of outliers. However, there are some applications for which these advantages are insufficient. Some kinds of data, such as loss distributions in insurance and traffic volumes on computer networks, involve distributions with very heavy tails, such that there may be doubts about whether even the first moment exists (see, e.g., Embrechts et al., 1997; Resnick, 1997; Willinger et al., 1998). For these applications it would be useful to have measures analogous to $L$-moments that remain meaningful for distributions that have no mean.

Trimmed $L$-moments, defined by Elamir and Seheult (2003), are generalizations of $L$-moments that do not require the mean of the underlying distribution to exist. They are defined by

$$\lambda_r^{(s,\,t)} = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathrm{E}(X_{r+s-j\,:\,r+s+t})\,. \tag{2}$$

Here $s$ and $t$ are positive integers. The case $s = t = 0$ yields the original $L$-moments defined by Hosking (1990). The term "trimmed" is appropriate because the definition of $\lambda_r^{(s,\,t)}$ does not involve the expectations of the $s$ smallest or the $t$ largest order statistics of the sample of size $r + s + t$. Trimmed $L$-moment ratios $\tau_r^{(s,\,t)} = \lambda_r^{(s,\,t)} / \lambda_2^{(s,\,t)}$ are dimensionless measures of the shape of a distribution.

Elamir and Seheult (2003) derived unbiased sample estimators of trimmed $L$-moments, gave expressions for the variances of these estimators, and obtained parameter estimates based on trimmed $L$-moments for several distributions. Karvanen (2005) used trimmed $L$-moments to estimate the parameters of a generalization of the Cauchy distribution. In this paper we give some further theoretical properties and illustrate some practical uses of trimmed $L$-moments.

In Section 2 we extend to trimmed $L$-moments some theoretical results that have been proved for $L$-moments: a sufficient condition for trimmed $L$-moments to exist, a proof that the trimmed $L$-moments uniquely define a distribution, a representation for the quantile function as a weighted function of orthogonal polynomials in which the coefficients are related to trimmed $L$-moments, and bounds on the numerical values of trimmed $L$-moment ratios. Section 3 illustrates some practical uses of trimmed $L$-moments: identification of distributions using a trimmed $L$-moment ratio diagram, and efficient estimation of the shape parameter of the generalized Pareto distribution. This shows how trimmed $L$-moments can be used similarly to $L$-moments, but continue to be useful for data drawn from heavy-tailed distributions. Section 4 describes an alternative way of defining trimmed $L$-moments that has some theoretical advantages. Section 5 cotains some concluding remarks. All proofs are deferred to Section 6.

# 2. Theoretical results concerning trimmed *L*-moments

## 2.1. Existence

Elamir and Seheult (2003) observe that a distribution may be specified by its trimmed $L$-moments even if some of its $L$-moments do not exist, but do not give explicit conditions under which this can occur. The following theorem gives sufficient conditions for the existence of trimmed $L$-moments.

**Theorem 1.** Let $X$ be a real-valued random variable. If $\mathrm{E}[\{\max(-X,0)\}^{1/(s+1)}]$ and $\mathrm{E}[\{\max(X,0)\}^{1/(t+1)}]$ exist, then so do all of the trimmed $L$-moments $\lambda_r^{(s,t)}$, $r=1,2,\ldots$, of $X$.

Thus for example the trimmed $L$-moments $\lambda_r^{(1,1)}$ exist if $E|X|^{1/2}<\infty$. This will be sufficient to permit the analysis of distributions describing many types of heavy-tailed data.

## 2.2. Uniqueness

Hosking (1990) showed that only one distribution can have a given set of $L$-moments. An analogous result is true for trimmed $L$-moments.

**Theorem 2.** A distribution for which the trimmed $L$-moments $\lambda_r^{(s,t)}$, $r=1,2,\ldots$, exist is characterized by those trimmed $L$-moments.

## 2.3. Relation to orthogonal polynomials

The $L$-moments of a random variable with quantile function $Q(\,.\,)$ have the representation

$$\lambda_{r+1} = \int_0^1 P_r^*(u)\,Q(u)\,du \tag{3}$$

(Hosking, 1990), where $P_r^*(u)$ is a shifted Legendre polynomial. These polynomials are orthogonal on the interval $[0,1]$ with constant weight function. An analogous result for trimmed $L$-moments involves polynomials that are orthogonal on the interval $[0,1]$ with weight function $u^s(1-u)^t$. These are shifted Jacobi polynomials, which we denote by

$$P_r^{*(t,s)}(x) = \sum_{j=0}^{r}(-1)^{r-j}\binom{r+t}{j}\binom{r+s}{r-j}u^j(1-u)^{r-j}\,. \tag{4}$$

Note that $P_r^{*(t,s)}(u) = P_r^{(t,s)}(2u-1)$, where $P_r^{(t,s)}(x)$ is the $r$th (unshifted) Jacobi polynomial, as defined for example by Abramowitz and Stegun (1972, eq. 22.2.1). The analogous result to (3) for trimmed $L$-moments is

$$\lambda_{r+1}^{(s,t)} = \frac{r!\,(r+s+t+1)!}{(r+1)\,(r+s)!\,(r+t)!} \int_0^1 u^s (1-u)^t \, P_r^{*(t,s)}(u)\, Q(u)\, du \, . \tag{5}$$

Integration by parts (for details see Section 6) gives a representation involving the derivative of the quantile function:

$$\lambda_{r+1}^{(s,t)} = \frac{(r-1)!\,(r+s+t+1)!}{(r+1)\,(r+s)!\,(r+t)!} \int_0^1 u^{s+1} (1-u)^{t+1} \, P_{r-1}^{*(t+1,s+1)}(u)\, Q'(u)\, du \, . \tag{6}$$

The substitution $u = F(x)$, where $F(x)$ is the cumulative distribution function, gives

$$\lambda_{r+1}^{(s,t)} = \frac{(r-1)!\,(r+s+t+1)!}{(r+1)\,(r+s)!\,(r+t)!} \int_0^1 \{F(x)\}^{s+1} \{1-F(x)\})^{t+1} \, P_{r-1}^{*(t+1,s+1)}(F(x))\, dx \, , \tag{7}$$

a form that does not require the existence of the derivative of the quantile function. Further integrations by parts give, for each $k = 0, 1, 2, \ldots, r$,

$$\lambda_{r+1}^{(s,t)} = \frac{(r-k)!\,(r+s+t+1)!}{(r+1)\,(r+s)!\,(r+t)!} \int_0^1 u^{s+k} (1-u)^{t+k} \, P_{r-k}^{*(t+k,s+k)}(u)\, Q^{(k)}(u)\, du \, , \tag{8}$$

provided that the required derivatives of the quantile function exist. This result has not previously been published, even for the original $L$-moments, for which it takes the simpler form

$$\lambda_{r+1} = \frac{(r-k)!}{r!} \int_0^1 u^k (1-u)^k \, P_{r-k}^{*(k,k)}(u)\, Q^{(k)}(u)\, du \, , \tag{9}$$

and for which the case $k = r$ gives the particularly striking result

$$\lambda_{r+1} = \frac{1}{r!} \int_0^1 u^r (1-u)^r \, Q^{(r)}(u)\, du \, . \tag{10}$$

$L$-moments can be defined as the coefficients in the expansion of the quantile function as a weighted sum of shifted Legendre polynomials Again, an analogous result for trimmed $L$-moments involves shifted Jacobi polynomials.

**Theorem 3.** Let $X$ be a continuous real-valued random variable such that $\mathrm{E}[\{\max(-X,0)\}^{2/(s+1)}]$ and $\mathrm{E}[\{\max(X,0)\}^{2/(t+1)}]$ are finite. Let $X$ have quantile function $Q(u)$ and trimmed $L$-moments $\lambda_r^{(s,\,t)}$, $r=1,2,\ldots$. Then the representation

$$Q(u) = \sum_{r=0}^{\infty} \frac{r(2r+s+t+1)}{r+s+t} \, \lambda_{r+1}^{(s,\,t)} \, P_r^{*(t,s)}(u) \qquad (11)$$

is convergent in the weighted mean square sense, with weight function $u^s(1-u)^t$.

## 2.4. Recurrences between trimmed *L*-moments with different degrees of trimming

Shifted Jacobi polynomials satisfy certain recurrence relations that can be used to derive corresponding relations between trimmed $L$-moments with different degrees of trimming. The main results are

$$(2r+s+t-1)\lambda_r^{(s,\,t)} = (r+s+t)\lambda_r^{(s,\,t-1)} - \frac{1}{r}(r+1)(r+s)\lambda_{r+1}^{(s,\,t-1)} \, , \qquad (12)$$

$$(2r+s+t-1)\lambda_r^{(s,\,t)} = (r+s+t)\lambda_r^{(s-1,\,t)} - \frac{1}{r}(r+1)(r+t)\lambda_{r+1}^{(s-1,\,t)} \, . \qquad (13)$$

For low orders of trimming successive applications of (12) and (13) yield the following relations between trimmed $L$-moments and $L$-moments:

$$\lambda_r^{(0,1)} = \frac{r+1}{2r}(\lambda_r - \lambda_{r+1}) \, ;$$

$$\lambda_r^{(0,2)} = \frac{(r+1)(r+2)}{2r(2r+1)}\lambda_r - \frac{r+2}{2r}\lambda_{r+1} + \frac{r+2}{2(2r+1)}\lambda_{r+2} \, ;$$

$$\lambda_r^{(1,1)} = \frac{(r+1)(r+2)}{2r(2r+1)}(\lambda_r - \lambda_{r+2}) \, .$$

These results are valid whenever all of the relevant trimmed $L$-moments exist. They are of mostly mathematical interest since trimmed $L$-moments such as $\lambda_r^{(1,1)}$ are likely to be of most use when $\lambda_r$ does not exist. However, the results do suggest ways of refining $L$-moment-based shape measures when they are near their bounds: for example when $\tau_3$ is close to 1, the measure $(\tau_3 - \tau_4)/(1-\tau_3)$ may be a useful additional measure of skewness, since it is a scalar multiple of the trimmed $L$-skewness $\tau_3^{(0,1)}$.

### 2.5. Bounds on trimmed *L*-moment ratios

The *L*-moment ratios $\tau_r$ have the convenient property that they are all bounded by 1 in absolute value. Trimmed *L*-moment ratios do not have this property. Instead we can show that

$$|\tau_r^{(s,t)}| \le \frac{2\,(m+1)!\,(r+s+t)!}{r\,(m+r-1)!\,(2+s+t)!} \qquad \text{where } m=\min(s,t). \tag{14}$$

Unless $s=t=0$, this bound is greater than 1 for all $r>2$ and increases as $r$ increases.

It is not difficult to find distributions for which $|\tau_r^{(s,t)}|>1$ for some $r$. For example consider the generalized Pareto distribution, with quantile function $Q(u)=\alpha\{1-(1-u)^k\}/k$. The skewness measure $\tau_3^{(1,1)}$ is a function of the shape parameter $k$ and exists for $k>-2$. We find that $\tau_3^{(1,1)}=10(1-k)/\{9(5+k)\}$, which exceeds 1 when $k<-35/19=-1.8421$ and approaches 10/9, the upper bound in (14), as $k\to-2$.

That the upper bound is not 1 makes trimmed *L*-moment ratios a little more difficult to interpret than the original *L*-moment ratios but is not a major practical problem. It could be avoided by changing the definition of $\lambda_r^{(s,t)}$ to include an appropriate rescaling, but this would make the definition (2) more complex and does not seem worthwhile.

## 3. Practical uses of trimmed *L*-moments

### 3.1. Sample trimmed *L*-moments

From an ordered data sample $x_1\le x_2\le\cdots\le x_n$, the trimmed *L*-moment $\lambda_r^{(s,t)}$ can be estimated unbiasedly by the "sample trimmed *L*-moment"

$$\ell_r^{(s,t)} = \frac{1}{r\binom{n}{r+s+t}} \sum_{j=s+1}^{n-t}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\binom{j-1}{r+s-k-1}\binom{n-j}{t+k}\,x_j \tag{15}$$

— this is a trivial generalization of Elamir and Seheult (2003, eq. (16)). The sample trimmed *L*-moment ratios $t_r^{(s,t)}=\ell_r^{(s,t)}/\ell_2^{(s,t)}$ are dimensionless measures of the shape of a data set, and are consistent estimators of the respective population quantities $\tau_r^{(s,t)}$.

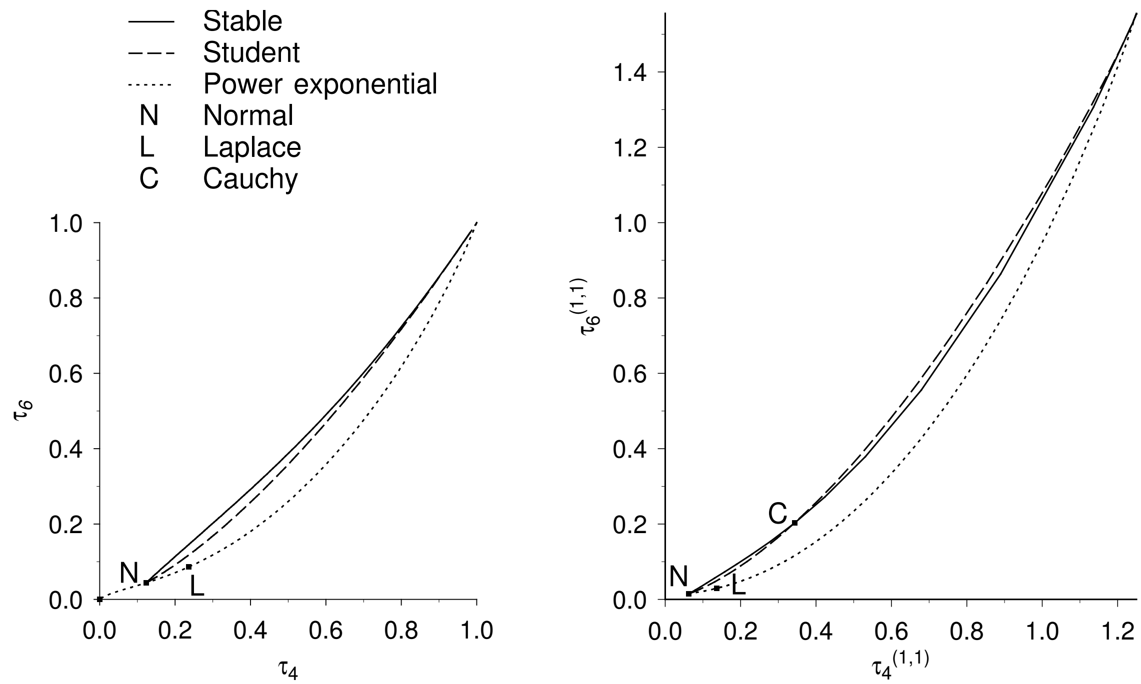


**Figure 1.** *L*-moment ratio diagrams for untrimmed and trimmed *L*-moment ratios of orders 4 and 6. Curves show the *L*-moment ratios of distributions in the indicated families, each of which has a single shape parameter. Marked points correspond to individual distributions.

### 3.2. The trimmed *L*-moment ratio diagram

An *L*-moment ratio diagram is a graph whose axes represent *L*-moments of different orders; parametric families of distributions occupy points, lines or regions on the graph and the sample *L*-moments of data sets can be plotted as points. This facilitates the comparison of data samples and the choice of which distribution to fit to a given data set. Hosking and Wallis (1997) give several examples. A similar diagram can be used with trimmed *L*-moments.

As an illustration we consider an *L*-moment ratio diagram using the even-order *L*-moment ratios $\tau_4$ and $\tau_6$. This is useful for distinguishing between symmetric distributions, as might arise in financial applications such as Value at Risk (e.g., Hosking et al., 2000). Figure 1 shows *L*-moment ratio diagrams for the original *L*-moment ratios and trimmed *L*-moment ratios with $s = t = 1$. The curves indicate the relations between fourth- and sixth-order *L*-moment ratios for three families of distributions: the stable and Student $t$, which each have power-law tails, and the

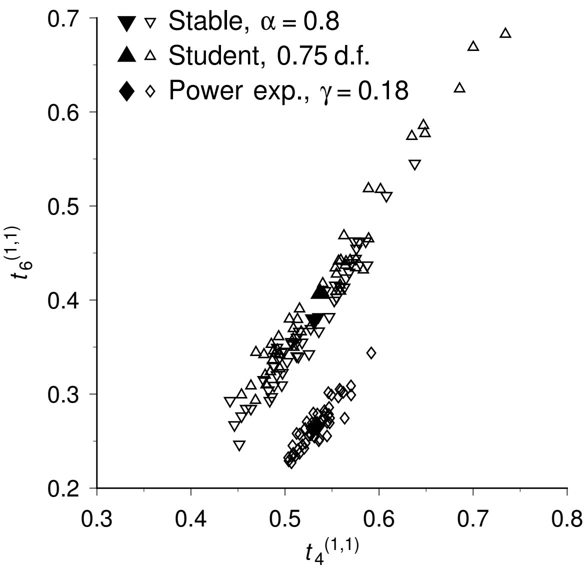


**Figure 2.** A trimmed $L$-moment ratio diagram for data samples. Solid symbols indicate the population trimmed $L$-moment ratios for the three indicated distibutions. Open symbols are sample trimmed $L$-moment ratios from 50 samples of size 1000 from each of the three distributions.

power exponential distribution, which has probability density function proportional to $\exp(-|x|^\gamma)$ and for which moments of all orders exist. Individual members of these families are shown as points. The general form of the diagrams is similar, but there are some important differences. The axis ranges are larger for the trimmed $L$-moment ratio diagram, reflecting the bounds of trimmed $L$-moment ratios derived in (14). The trimmed $L$-moment ratio diagram covers a wider range of distributions; for example the Cauchy distribution, which on the $L$-moment ratio diagram is the limiting point as $\tau_4 \to 1$ and $\tau_6 \to 1$, appears on the trimmed $L$-moment ratio diagram at the point $\tau_4^{(1,1)} = 0.342, \tau_6^{(1,1)} = 0.202$: points on the "Stable" and "Student" curves to the right of the Cauchy point correspond to distributions for which the mean does not exist.

Figure 2 illustrates the ability of trimmed $L$-moment ratios to distinguish between samples from heavy-tailed distributions. The large circles indicate the population trimmed $L$-moment ratios from each of three distributions: stable with shape parameter 0.8, Student $t$ with 0.75 degrees of freedom, and power exponential with exponent $\gamma = 0.18$. Each of these distributions has $\tau_4^{(1,1)}$ close to 0.53, but their values of $\tau_6^{(1,1)}$ differ. For these stable and Student $t$ distributions the mean does not exist. The plotted points are the sample trimmed $L$-moment ratios $t_4^{(1,1)}$ and $t_6^{(1,1)}$ from 50 samples of size 1000 from each of these three distributions. There is some overlap between the clouds of points corresponding to the stable and Student $t$ distributions, which have similarly heavy tails, but apart from this the sample trimmed $L$-moment ratios provide an effective way of distinguishing between these distributions.

### 3.3. Estimation for the generalized Pareto distribution

The generalized Pareto distribution has quantile function $Q(u) = \alpha\{1 - (1-u)^k\}/k$. Its upper tail has a power-law form, with probability density function $f(x) \sim x^{1/k-1}$ as $x \to \infty$. It is a widely used model for distributions with heavy tails in fields such as insurance, finance and network traffic modeling. The distribution may be fitted to an entire data set or to the excesses over some threshold. The mean of the distribution does not exist when $k \le -1$; in some applications these values of $k$ are plausible in practice. Estimation of the shape parameter, and hence of the tail weight, is often of most interest.

Because the distribution has only one infinite tail, the asymmetrically trimmed $L$-moments $\lambda_r^{(0,t)}$ seem likely to be of most use. We can compute these quantities using (2): for example

$$\lambda_1^{(0,t)} = \frac{\alpha}{1+t+k}, \qquad \lambda_2^{(0,t)} = \frac{(2+t)\alpha}{2(1+t+k)(2+t+k)}.$$

The parameters can be expressed as a function of trimmed $L$-moments of orders 1 and 2:

$$k = (t+2)(\tfrac{1}{2}\lambda_1^{(0,t)}/\lambda_2^{(0,t)} - 1), \qquad \alpha = (t+1+k)\lambda^{(0,t)}.$$

Replacing population trimmed $L$-moments by their sample estimates, we obtain an estimator of the shape parameter:

$$\hat{k}^{(0,t)} = (t+2)(\tfrac{1}{2}\ell_1^{(0,t)}/\ell_2^{(0,t)} - 1), \qquad \hat{\alpha}^{(0,t)} = (t+1+k)\ell^{(0,t)}. \tag{16}$$

Hosking and Wallis (1987) showed that probability weighted moment estimators (equivalent to estimators based on $L$-moments) of parameters and quantiles of the generalized Pareto distribution are generally more accurate than the maximum-likelihood estimates, for shape parameters in the range $-0.5 < k < 0$ and sample sizes up to 500.

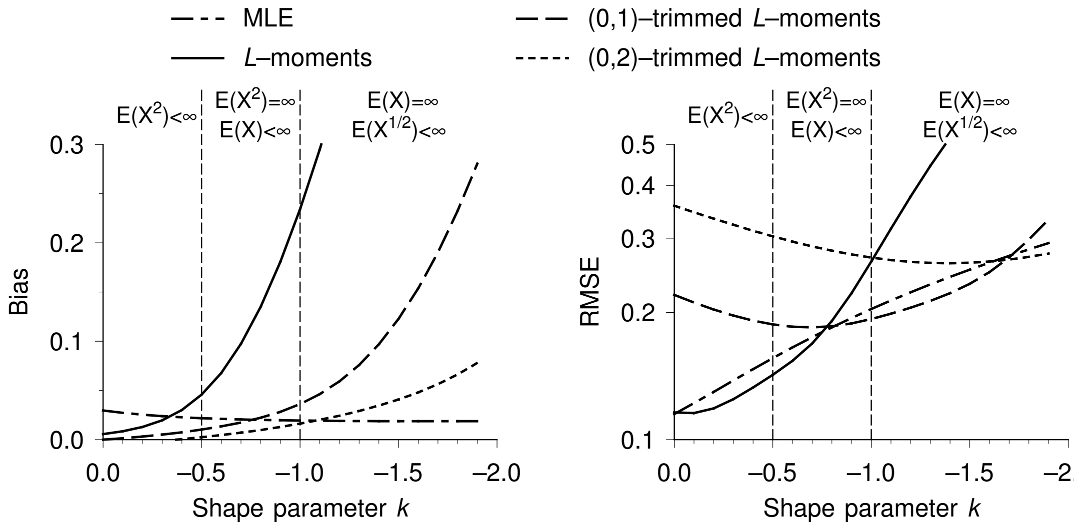


**Figure 3.** Bias and RMSE of estimates of the shape parameter of the generalized Pareto distribution for sample size 100.

For the maximum-likelihood estimate and trimmed $L$-moment estimators, bias and root mean square error (RMSE) of estimates of the shape parameter were obtained using Monte Carlo simulation. Figure 3 shows the results for sample size 100 over a range of values of $k$. As $k$ decreases and the tail weight of the distribution increases, larger amounts of trimming become preferable: first $\hat{k}^{(0,0)}$, then $\hat{k}^{(0,1)}$, then $\hat{k}^{(0,2)}$ has the lowest RMSE of the trimmed $L$-moment estimators. Each of the trimmed $L$-moment estimators has a range of $k$ values for which it has lower RMSE than the maximum-likelihood estimate. Thus, particularly if prior knowledge enables the user to judge the appropriate degree of trimming, trimmed $L$-moment estimators can outperform the maximum-likelihood estimate. Note that the trimmed $L$-moment estimators can be computed straightforwardly using (16), while the computation of the maximum-likelihood estimates requires iterative procedures.

### 3.4. Example: network traffic volumes

Table 1 shows a data set of 720 hourly measurements of traffic through one node of a computer network. This is a kind of data for which heavy-tailed distributions are often appropriate. Generalized Pareto distributions were fitted using $L$-moments, trimmed $L$-moments, and the method of maximum likelihood, to excess over various thresholds. Estimates of the shape parameter, which determines the tail weight, are

**Table 1.** Network traffic data.

Smaller values, shown as their frequency in the data set:

| | _0 | _1 | _2 | _3 | _4 | _5 | _6 | _7 | _8 | _9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0–9 | 288 | 105 | 60 | 18 | 17 | 17 | 8 | 14 | 8 | 4 |
| 10–19 | 6 | 10 | 4 | 2 | 7 | 5 | 3 | 3 | 3 | 1 |
| 20–29 | 5 | 5 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 2 |
| 30–39 | 1 | 5 | 2 | 1 | 1 | 0 | 0 | 1 | 2 | 1 |
| 40–49 | 3 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |

Larger values, listed individually:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 50 | 53 | 54 | 56 | 57 | 59 | 61 | 62 | 63 | 65 |
| 69 | 70 | 76 | 76 | 79 | 83 | 83 | 86 | 90 | 91 |
| 95 | 95 | 105 | 107 | 108 | 122 | 123 | 125 | 130 | 134 |
| 134 | 146 | 159 | 161 | 181 | 183 | 209 | 209 | 231 | 238 |
| 250 | 253 | 274 | 274 | 282 | 285 | 297 | 303 | 328 | 335 |
| 345 | 352 | 356 | 371 | 374 | 423 | 484 | 521 | 526 | 584 |
| 590 | 616 | 642 | 663 | 706 | 706 | 745 | 769 | 769 | 777 |
| 794 | 903 | 953 | 1086 | 1225 | 1383 | 1462 | 1512 | 1709 | 1768 |
| 1808 | 3134 | 5955 | 6405 | 20480 | | | | | |

**Table 2.** Shape parameter estimates for excesses over various thresholds for the data in Table 1. $n$ is the number of data values that exceed the given threshold.

| | | Shape parameter estimates | | | |
|---|---|---|---|---|---|
| Threshold | $n$ | $\hat{k}^{(0,0)}$ | $\hat{k}^{(0,1)}$ | $\hat{k}^{(0,2)}$ | $\hat{k}_{\text{MLE}}$ |
| 0 | 432 | –0.92 | –1.56 | –1.75 | –1.74 |
| 10 | 175 | –0.83 | –1.34 | –1.73 | –1.56 |
| 20 | 132 | –0.79 | –1.24 | –1.69 | –1.45 |
| 50 | 84 | –0.71 | –0.89 | –1.18 | –0.88 |
| 100 | 63 | –0.67 | –0.64 | –0.74 | –0.69 |
| 200 | 49 | –0.68 | –0.62 | –0.61 | –0.71 |
| 500 | 28 | –0.72 | –0.72 | –0.44 | –0.82 |

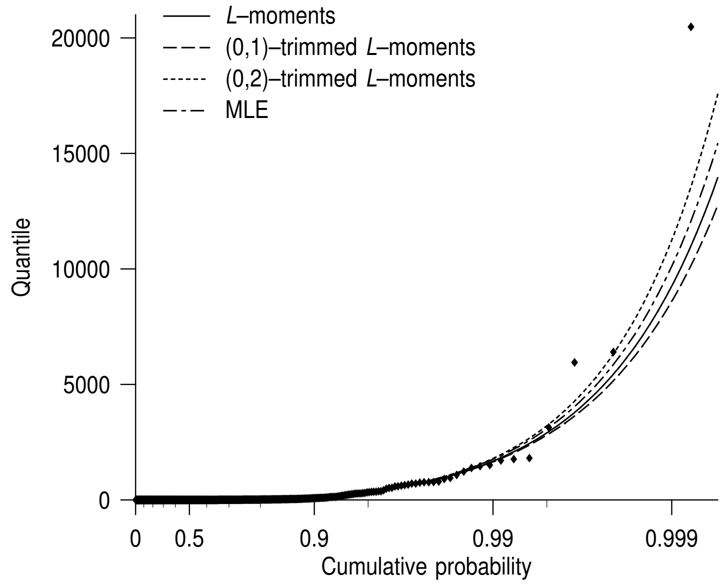


**Figure 4.** Data from Table 1 and distributions fitted to data values in excess of the threshold 100.

given in Table 2. As the threshold increases the estimated shape parameter generally takes smaller negative values. This pattern is less extreme for the untrimmed estimator $\hat{k}^{(0,0)}$, which cannot take values less than –1. However, for thresholds in the 100–200 range, the estimates for all methods are in the same range, –0.6 to –0.7, suggesting that data values above these values are well described by a generalized Pareto distribution. The fitted distributions for threshold 100 are illustrated in Figure 4.

## 4. Alternative trimmed *L*-moments

An alternative definition of trimmed *L*-moments can be derived from the representation for *L*-moments based on (6), i.e.

$$\lambda_{r+1} = \frac{1}{r} \int_0^1 u(1-u)\, P^{*(1,1)}_{r-1}(u)\, Q'(u)\, du \,. \tag{17}$$

We define alternative trimmed *L*-moments by $\tilde{\lambda}_1^{(s,\,t)} = \mathrm{E}\, X_{s+1:s+t+1}$ and, for a distribution with differentiable quantile function $Q(u)$,

$$\tilde{\lambda}_{r+1}^{(s,\,t)} = \frac{(s+t+1)!}{(r-1)\, s!\, (t+1)!} \int_0^1 u^{s+1}(1-u)^{t+1}\, P^{*(1,1)}_{r-1}(u)\, Q'(u)\, du\,, \qquad r = 2, 3, \ldots. \tag{18}$$

The constant multiplying the integral in (18) is chosen so that the analog of the coefficient of variation for trimmed $L$-moments, $\tilde{\lambda}_2^{(s,t)}/\tilde{\lambda}_1^{(s,t)}$, behaves like its untrimmed equivalent $\lambda_2/\lambda_1$ in that it takes values between 0 and 1 for positive random variables.

It is clear from (18) that $\tilde{\lambda}_r^{(s,t)}$ is the $r$th (untrimmed) $L$-moment of a distribution with quantile function derivative $u^s(1-u)^t Q'(u)$. These trimmed $L$-moments therefore have the same range of feasible values as $L$-moments; in particular the $L$-moment ratios $\tilde{\tau}_r^{(s,t)} = \tilde{\lambda}_r^{(s,t)}/\tilde{\lambda}_2^{(s,t)}$ are all bounded by 1 in absolute value. As a further example, because the lower bound for $\tau_4$ given $\tau_3$ is $\tau_4 \ge (5\tau_3^2 - 1)/4$ (Hosking, 1990, eq. (2.7)) it immediately follows that the lower bound for $\tilde{\tau}_4^{(s,t)}$ given $\tilde{\tau}_3^{(s,t)}$ is $\tilde{\tau}_4^{(s,t)} \ge \{5(\tilde{\tau}_3^{(s,t)})^2 - 1\}/4$; this is not true for the trimmed $L$-moments defined by (2).

Reversing the integration by parts that led to (6), we obtain an expression for $\tilde{\lambda}_r^{(s,t)}$ in terms of expected order statistics:

$$\tilde{\lambda}_r^{(s,t)} = \frac{1}{r-1} \sum_{k=0}^{r-2} \frac{\binom{r-1}{k}\binom{r-1}{k+1}}{\binom{r+s+t}{t+1+k}} \left( \mathrm{E}\, X_{r+s-k:r+s+t} - \mathrm{E}\, X_{r+s-k-1:r+s+t} \right), \quad r \ge 2. \tag{19}$$

This definition is valid even for distributions for which the quantile function is not differentiable. It is of course not as simple as (2), but can be evaluated straightforwardly when needed. It can be used as the basis for deriving unbiased sample estimates of the alternative trimmed $L$-moments, by the same procedure as for trimmed $L$-moments (Elamir and Seheult, 2003, sec. 3).

In the case $s = t = 1$ we obtain the following alternative trimmed $L$-moments:

$$\tilde{\lambda}_1^{(1,1)} = \mathrm{E}\, X_{2:3},$$
$$\tilde{\lambda}_2^{(1,1)} = \tfrac{1}{6}\, \mathrm{E}(X_{3:4} - X_{2:4}),$$
$$\tilde{\lambda}_3^{(1,1)} = \tfrac{1}{10}\, \mathrm{E}(X_{4:5} - 2X_{3:5} + X_{2:5}),$$
$$\tilde{\lambda}_4^{(1,1)} = \tfrac{1}{60}\, \mathrm{E}(4X_{5:6} - 13X_{4:6} + 13X_{3:6} - 4X_{2:6}).$$

These are still linear combinations of expected order statistics, but from order 4 onwards the relative magnitudes of the coefficients of the linear combinations differ from those of the trimmed $L$-moments defined in (2). An interesting point is that the alternative trimmed $L$-moments $\tilde{\lambda}_r^{(s,t)}$ are zero for $r \geq 3$ for a distribution with quantile function derivative proportional to $u^{-s}(1-u)^{-t}$, because the polynomial in the integrand of (18) is then proportional to $u(1-u)$, the weight function of the orthogonal polynomials $P_r^{*(1,1)}(u)$. For $s=t=1$ this distribution is the logistic, with quantile function $Q(u)=\alpha \log(u/(1-u))$. Thus the alternative trimmed $L$-moments $\tilde{\lambda}_r^{(1,1)}$ can be regarded as $L$-moments that have been centered on the logistic distribution, in contrast with the untrimmed $L$-moments which are centered on the uniform distribution.

## 5. Concluding remarks

Trimmed $L$-moments are a natural extension of the $L$-moments defined by Hosking (1990). Several of their mathematical properties are shared with, or are extensions of, those of $L$-moments, and they can used in practice in a similar way to $L$-moments. They provide simple and effective methods of inference for distribution with tails so heavy that the mean does not exist.

A question not discussed in this paper is how to decide, other than by prior information, on which degrees of trimming are appropriate when using trimmed $L$-moments. In some cases it may be possible to decide this question based on the data themselves. The resulting statistics might be termed "adaptively trimmed $L$-moments". They will be discussed in a subsequent paper.

## 6. Proofs

**Proof of Theorem 1.** The conditions $\mathrm{E}[\{\max(-X,0)\}^{1/(s+1)}]$ and $\mathrm{E}[\{\max(X,0)\}^{1/(t+1)}]$ ensure the existence of the integrals $\int_0 u^s Q(u)du$ and $\int^1 (1-u)^t Q(u)du$ respectively. To see this, first consider the integral $\int_u^1 (1-v)^t Q(v)dv$ for some $u$ with $0<u<1$; without loss of generality take $u$ such that $Q(u)>0$ (if there is no such $u$ then the integral is certainly finite). Let $Z(v)=\{Q(v)\}^{1/(t+1)}$; note that $Z(v)$ is an increasing function of $v$ and that, by the assumptions of the theorem, $\int_u^1 Z(v)dv < \infty$. We have

$$
\begin{aligned}
\int_u^1 (1-v)^t\, Q(v)\, dv &= \int_u^1 (1-v)^t \{Z(v)\}^{t+1}\, dv \\
&= t! \underset{u<v_1<v_2<\cdots<v_{t+1}<1}{\int\int\cdots\int} \{Z(v_1)\}^{t+1}\, dv_1\, dv_2\, \ldots\, dv_{t+1} \\
&\le t! \underset{u<v_1<v_2<\cdots<v_{t+1}<1}{\int\int\cdots\int} Z(v_1)\, Z(v_2)\, \ldots\, Z(v_{t+1})\, dv_1\, dv_2\, \ldots\, dv_{t+1} \\
&\le t! \int_u^1 \int_u^1 \cdots \int_u^1 Z(v_1)\, Z(v_2)\, \ldots\, Z(v_{t+1})\, dv_1\, dv_2\, \ldots\, dv_{t+1} \\
&= t! \left( \int_u^1 Z(v)\, dv \right)^{t+1} \\
&< \infty\,.
\end{aligned}
$$

A similar argument shows that, for $0<u<1$, $\int_0^u v^s\, Q(v)\, dv$ is finite. Therefore so is $\int_0^1 u^s(1-u)^t\, Q(u)\, du$, which apart from a multiplicative constant is $\mathrm{E}\, X_{s+1:s+t+1}$ (David and Nagaraja, 2003, eq. (3.1.1′)). Thus, under the conditions of the theorem, $\lambda_1^{(s,\,t)} = \mathrm{E}\, X_{s+1:s+t+1}$ exists.

From (2), the higher-order trimmed $L$-moments $\lambda_r^{(s,\,t)}$, $r>1$, are sums of expectation of order statistics $X_{j:r+s+t}$ with $j>s$ or $j \le r+s$, which can be written as multiples of $\int_0^1 u^S(1-u)^T\, Q(u)\, du$ with $S \ge s$ and $T \ge t$ and by the same argument as above are all finite; thus each $\lambda_r^{(s,\,t)}$ exists, for $r=1, 2, \ldots$.

**Proof of Theorem 2.** From (2), $\lambda_r^{(s,\,t)}$ is a sum of expectation of order statistics $X_{j:r+s+t}$ with $j>s$ or $j \le r+s$, each of which which can be written as a multiple of $\int_0^1 u^S(1-u)^T\, Q(u)\, du$ with $S \ge s$ and $T \ge t$ and expressed as a weighted sum of integrals of the form $\int_0^1 u^s(1-u)^k\, Q(u)\, du$, $k=t, t+1, \ldots, r+t-s$, i.e. as a weighted sum of expected values of order statistics $\mathrm{E}\, X_{s+1:s+t+k}$, $k=1, 2, \ldots, r$. Thus from the trimmed $L$-moments $\lambda_r^{(s,\,t)}$, $r=1, \ldots, R$, one can determine the expected order statistics $\mathrm{E}\, X_{s+1:s+t+k}$, $k=1, 2, \ldots, R$, and from the complete set of trimmed $L$-moments $\lambda_r^{(s,\,t)}$, $r=1, 2, \ldots$, one can determine the infinite sequence of expected order statistics $\mathrm{E}\, X_{s+1:s+t+k}$, $k=1, 2, \ldots$. This sequence of expected order statistics suffices to determine the distribution (Huang, 1975). Therefore the set of trimmed $L$-moments $\lambda_r^{(s,\,t)}$, $r=1, 2, \ldots$, also determines the distribution.

**Proof of (5).** The proof is straightforward: it involves using the expression

$$\mathrm{E}\, X_{r:n} = \frac{n!}{(r-1)!\,(n-r)!} \int_0^1 u^{r-1}(1-u)^{n-r}\, Q(u)\, du \tag{20}$$

(David and Nagaraja, 2003, eq. (3.1.1′)) to substitute for the expected values in (2), expressing the resulting sum of integrals as the integral of a sum of powers of $u^{r+s-k}(1-u)^{t+k}$, and verifying that the integral in (5) has the same form.

**Proof of (6).** From the Rodrigues formula for Jacobi polynomials (e.g., Abramowitz and Stegun, 1972, eq. 22.11.1, for the unshifted polynomials) we have

$$\frac{d}{du}\{u^s(1-u)^t\, P_r^{*(t,s)}(u)\} = -(r+1)u^{s-1}(1-u)^{t-1} P_{r+1}^{*(t-1,s-1)}(u)\,. \tag{21}$$

If the quantile function $Q$ is differentiable, we can integrate (5) by parts, using (21), to give

$$\int_0^1 u^s(1-u)^t\, P_r^{*(t,s)}(u)\, Q(u)\, du = \frac{1}{r}\int_0^1 u^{s+1}(1-u)^{t+1}\, P_{r-1}^{*(t+1,s+1)}(u)\, Q'(u)\, du \tag{22}$$

— the integrated term, $\Big[r^{-1}\int_0^1 u^{s+1}(1-u)^{t+1}\, P_{r-1}^{*(t+1,s+1)}(u)\, Q(u)\Big]_0^1$, is zero because existence of the trimmed $L$-moments ensures that $u^{s+1}Q(u)\to 0$ as $u\to 0$ and $(1-u)^{t+1}Q(u)\to 0$ as $u\to 1$.

**Proof of Theorem 3.** We seek an approximation to the quantile function $Q(u)$ of the form

$$Q(u) = \sum_{r=0}^{R} a_r\, P_r^{*(t,s)}(u), \qquad 0<u<1. \tag{23}$$

To determine the $a_r$ we denote the error of the approximation (23) by

$$E_R(u) = Q(u) - \sum_{r=0}^{R} a_r\, P_r^{*(t,s)}(u)$$

and seek to minimize the weighted mean square error $\int_0^1 u^s(1-u)^t\,\{E_R(u)\}^2\, du$. The moment conditions given in the Theorem ensure that this integral is finite, by a proof similar to that of Theorem 1. We have

$$\int_0^1 u^s (1-u)^t \{E_R(u)\}^2 \, du$$

$$= \int_0^1 u^s (1-u)^t \{Q(u)\}^2 \, du - 2 \sum_{r=0}^{R} a_r \int_0^1 u^s (1-u)^t \, P_r^{*(t,s)}(u) \, Q(u) \, du$$

$$+ \sum_{r=0}^{R} \sum_{v=0}^{R} a_r a_v \int_0^1 u^s (1-u)^t \, P_r^{*(t,s)}(u) \, P_v^{*(t,s)}(u) \, du$$

$$= \int_0^1 u^s (1-u)^t \{Q(u)\}^2 \, du - 2 \sum_{r=0}^{R} a_r \, \frac{(r+1)\,(r+s)!\,(r+t)!}{r!\,(r+s+t+1)} \, \lambda_{r+1}^{(s,\,t)}$$

$$+ \sum_{r=0}^{R} a_r^2 \, \frac{(r+s)!\,(r+t)!}{(2r+s+t+1) r! (r+s+t)!} \; ;$$

here we have used (5) and the orthogonality properties of the shifted Jacobi polynomials (see Abramowitz and Stegun, 1972, eq. 22.2.1, for the unshifted version). The weighted mean square error is minimized by choosing

$$a_r = \frac{r(2r+s+t+1)}{r+s+t} \, \lambda_{r+1}^{(s,\,t)} \, .$$

That the weighted mean square error tends to zero as $R \to \infty$, i.e. that the set of shifted Jacobi polynomials is complete, is shown by Szegö (1959, chap. 4).

**Proof of (12) and (13).** The key results are recurrence relations for Jacobi polynomials (Abramowitz and Stegun, 1972, eqs. 22.7.15 and 22.7.16), which when expressed in terms of shifted Jacobi polynomials become

$$(2r+s+t+1)(1-u) P_r^{*(t,s)}(u) = (r+t) P_r^{*(t-1,s)}(u) - (r+1) P_{r-1}^{*(t-1,s)}(u) \, ,$$

$$(2r+s+t+1) u P_r^{*(t,s)}(u) = (r+s) P_r^{*(t,s-1)}(u) + (r+1) P_{r-1}^{*(t,s-1)}(u) \, .$$

Substituting these expressions into (5) immediately yields (12) and (13).

**Proof of (14).** Using (6) with the substitution $u = F(x)$ we obtain

$$\lambda_{r+1}^{(s,\,t)} = c_r \int_0^1 \{F(x)\}^{s+1} \{1 - F(x)\}^{t+1} \, P_{r-1}^{*(t+1,s+1)}(F(x)) \, dx \qquad (24)$$

where

$$c_r = \frac{(r-1)!\,(r+s+t+1)!}{(r+1)\,(r+s)!\,(r+t)!}\,. \tag{25}$$

Now

$$\max_{0\le u\le 1} |P^{*(t+1,s+1)}_{r-1}| = \binom{r+M}{r-1} \qquad \text{where } M = \max(s,t) \tag{26}$$

(Abramowitz and Stegun, 1972, eq. 22.14.1), and $\{F(x)\}^{s+1}\{1-F(x)\}^{t+1} \ge 0$ for all $x$ with $0 < x < 1$, so

$$|\lambda^{(s,t)}_{r+1}| \le c_r \binom{r+M}{r-1} \int_0^1 \{F(x)\}^{s+1}\{1-F(x)\}^{t+1}\,dx = c_r \binom{r+M}{r-1} \frac{1}{c_1}\lambda^{(s,t)}_2\,.$$

Thus, writing $m = \min(s,t)$, we have

$$\begin{aligned}
|\tau^{(s,t)}_{r+1}| &= |\lambda^{(s,t)}_{r+1}|/\lambda^{(s,t)}_2 \\
&\le \frac{c_r}{c_1}\binom{r+M}{r-1} \\
&= \frac{(r-1)!\,(r+s+t+1)!}{(r+1)\,(r+m)!\,(r+M)!}\;\frac{2\,(2+m)!\,(2+M)!}{(2+s+t)!}\;\frac{(r+M)!}{(r-1)!\,(M+1)!} \\
&= \frac{2\,(m+1)!\,(r+s+t)!}{r\,(m+r-1)!\,(2+s+t)!}\,.
\end{aligned}$$

# References


Abramowitz, M., and Stegun, I. A. (1972). *Handbook of mathematical functions.* Washington, D.C.: National Bureau of Standards.

David, H. A., and Nagaraja, H. (2003). *Order Statistics*, 3rd edn. New York: Wiley.

Elamir, E. A. H., and Seheult, A. H. (2003). Trimmed $L$-moments. *Comput. Statist. Data Anal.*, **43**, 299–314.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events.* Berlin: Springer-Verlag.

Hosking, J. R. M. (1990). $L$-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc. B*, 52, 105–124.

Hosking, J., Bonti, G., and Siegel, D. (2000). Beyond the lognormal, *Risk*, **13:5**, 59–62.

Hosking, J. R. M., and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29, 339–349.

Hosking, J. R. M., and Wallis, J. R. (1997), *Regional frequency analysis: an approach based on L-moments*, Cambridge, England: Cambridge University Press.

Huang, J. S. (1975). Characterization of distributions by expected values of the order statistics. *Ann. Inst. Statist. Math.*, **27**, 87–93.

Karvanen, J. (2005). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Comput. Statist. Data Anal.*, forthcoming. Preprint at `http://www.bsp.brain.riken.jp/publications/2005/karvanen_quantile_mixtures.pdf`.

Resnick, S. I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.*, **25**, 1805–1869.

Szegö, G. (1959). *Orthogonal polynomials*, 2nd ed. New York: American Mathematical Society.

Willinger, W., Paxson, V., and Taqqu, M. S. (1998). Self-similarity and heavy tails: structural modeling of network traffic. In *A Practical Guide to Heavy Tails*, eds. Adler, R., Feldman, R., and Taqqu, M. S., pp. 27–53. Boston: Birkhäuser.